NATIONAL QUALITY FORUM

National Consensus Standards for Behavioral Health Conditions

Standing Committee Orientation

Tracy Lustig
Karen Johnson
Kirsten Reed
Desmirra Quinnonez

January 19, 2017

Welcome

Project Team

- Tracy Lustig, DPM, MPH
  - *Senior Director*
- Kirsten Reed
  - *Project Manager*
- Desmirra Quinnonez
  - *Project Analyst*

Agenda for the Call

- Standing Committee Introductions
- Overview of NQF, the Consensus Development Process, and Roles of the Standing Committee, co-chairs, NQF staff
- Overview of NQF's portfolio of behavioral health measures
- Review of project activities and timelines
- Overview of NQF's measure evaluation criteria
- Overview of SDS Trial Period
- SharePoint Tutorial
- Next steps

Behavioral Health Standing Committee

- Peter Briss, MD, MPH, (Co-Chair)
- Harold Pincus, MD (Co-Chair)
- Robert Atkins, MD, MPH
- Mady Chalk, PhD, MSW
- Shane Coleman, MD, MPH*
- David Einzig, MD
- Julie Goldstein Grumet, PhD
- Charles Gross, PhD*
- Constance Horgan, ScD
- Lisa Jensen, DNP, APRN
- Dolores (Dodi) Kelleher, MS, DMH
- Kraig Knudsen, PhD
- Michael R. Lardieri, LCSW
- Tami Mark, PhD, MBA
- Raquel Mazon Jeffers, MPH, MIA
- Bernadette Melnk, PhD, RN, CPNP/FAANP, FNAP, FAAN
- Laurence Miller, MD
- Brooke Parish, MD*
- David Pating, MD
- Vanita Pindolia, PharmD
- Rhonda Robinson Beale, MD
- Lisa Shea, MD, DFAPA
- Andrew Sperling, JD*
- Jeffery Susman, MD
- Michael Trangle, MD
- Bonnie Zima, MD, MPH
- Leslie S. Zun, MD, MBA

*indicates new committee member

Overview of NQF, the CDP, and Roles

The National Quality Forum: A Unique Role

Established in 1999, NQF is a non-profit, non-partisan, membership-based organization that brings together public and private sector stakeholders to reach consensus on healthcare performance measurement. The goal is to make healthcare in the U.S. better, safer, and more affordable.

Mission: To lead national collaboration to improve health and healthcare quality through measurement

- An Essential Forum
- Gold Standard for Quality Measurement
- Leadership in Quality

NQF Activities in Multiple Measurement Areas

- **Performance Measure Endorsement**
  - *600+ NQF-endorsed measures across multiple clinical areas*
  - *19 empaneled standing committees*
- **Measure Applications Partnership (MAP)**
  - *Advises HHS on selecting measures for 20+ federal programs, Medicaid, and health exchanges*
- **National Quality Partners**
  - *Convenes stakeholders around critical health and healthcare topics*
  - *Spurs action on patient safety, early elective deliveries, and other issues*
- **Measurement Science**
  - *Convenes private and public sector leaders to reach consensus on complex issues in healthcare performance measurement such as attribution, alignment, sociodemographic status (SDS) adjustment*

NQF Consensus Development Process (CDP)

7 Steps for Measure Endorsement

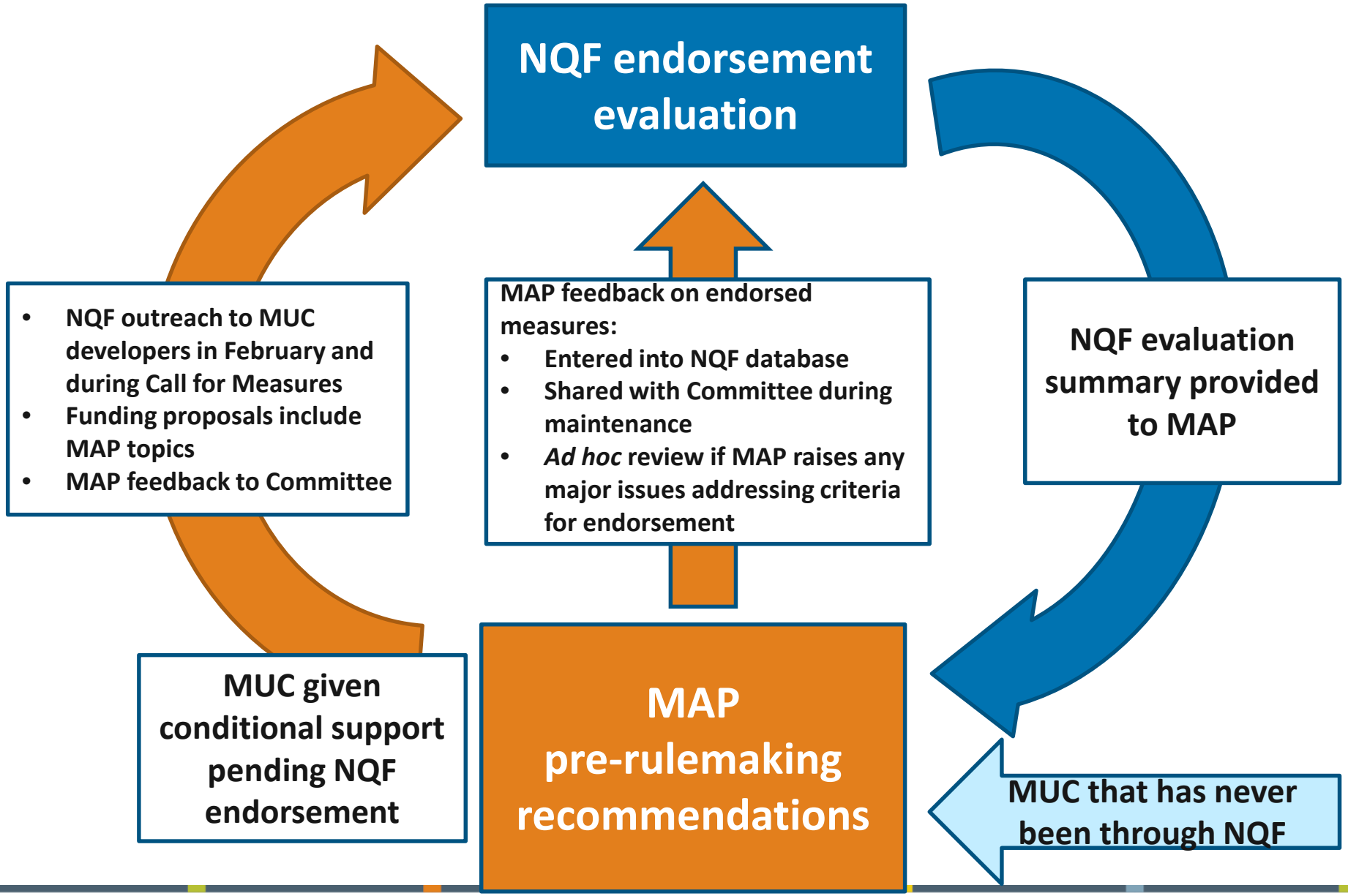
- Call for nominations for Standing Committee
- Call for candidate standards (measures)
- Candidate consensus standards review
- Public and member comment
- NQF member voting
- Consensus Standards Approval Committee (CSAC) ratification and endorsement
- Appeals

Measure Application Partnership (MAP)

In pursuit of the National Quality Strategy, the MAP:

- Informs the selection of performance measures to achieve the goal of improvement, transparency, and value for all
- Provides input to HHS during pre-rulemaking on the selection of performance measures for use in public reporting, performance-based payment, and other federal programs
- Identifies gaps for measure development, testing, and endorsement
- Encourages measurement alignment across public and private programs, settings, levels of analysis, and populations to:
  - *Promote coordination of care delivery*
  - *Reduce data collection burden*

CDP-MAP INTEGRATION – INFORMATION FLOW



Role of the Standing Committee
General Duties

- Act as a proxy for the NQF multi-stakeholder membership
- Serve 2-year or 3-year terms
- Work with NQF staff to achieve the goals of the project
- Evaluate candidate measures against the measure evaluation criteria
- Respond to comments submitted during the review period
- Respond to any directions from the CSAC

Role of the Standing Committee
Measure Evaluation Duties

- All members review ALL measures
- Evaluate measures against each criterion
  - *Indicate the extent to which each criterion is met and rationale for the rating*
- Make recommendations to the NQF membership for endorsement
- Oversee behavioral health portfolio of measures
  - *Promote alignment and harmonization*
  - *Identify gaps*

Role of the Standing Committee Co-Chairs

- Co-facilitate Standing Committee (SC) meetings
- Work with NQF staff to achieve the goals of the project
- Assist NQF in anticipating questions and identifying additional information that may be useful to the SC
- Keep SC on track to meet goals of the project without hindering critical discussion/input
- Represent the SC at CSAC meetings
- Participate as a SC member

Role of NQF Staff

- NQF project staff works with SC to achieve the goals of the project and ensure adherence to the consensus development process:
  - *Organize and staff SC meetings and conference calls*
  - *Guide the SC through the steps of the CDP and advise on NQF policy and procedures*
  - *Review measure submissions and prepare materials for Committee review*
  - *Draft and edit reports for SC review*
  - *Ensure communication among all project participants (including SC and measure developers)*
  - *Facilitate necessary communication and collaboration between different NQF projects*

Role of NQF Staff
Communication

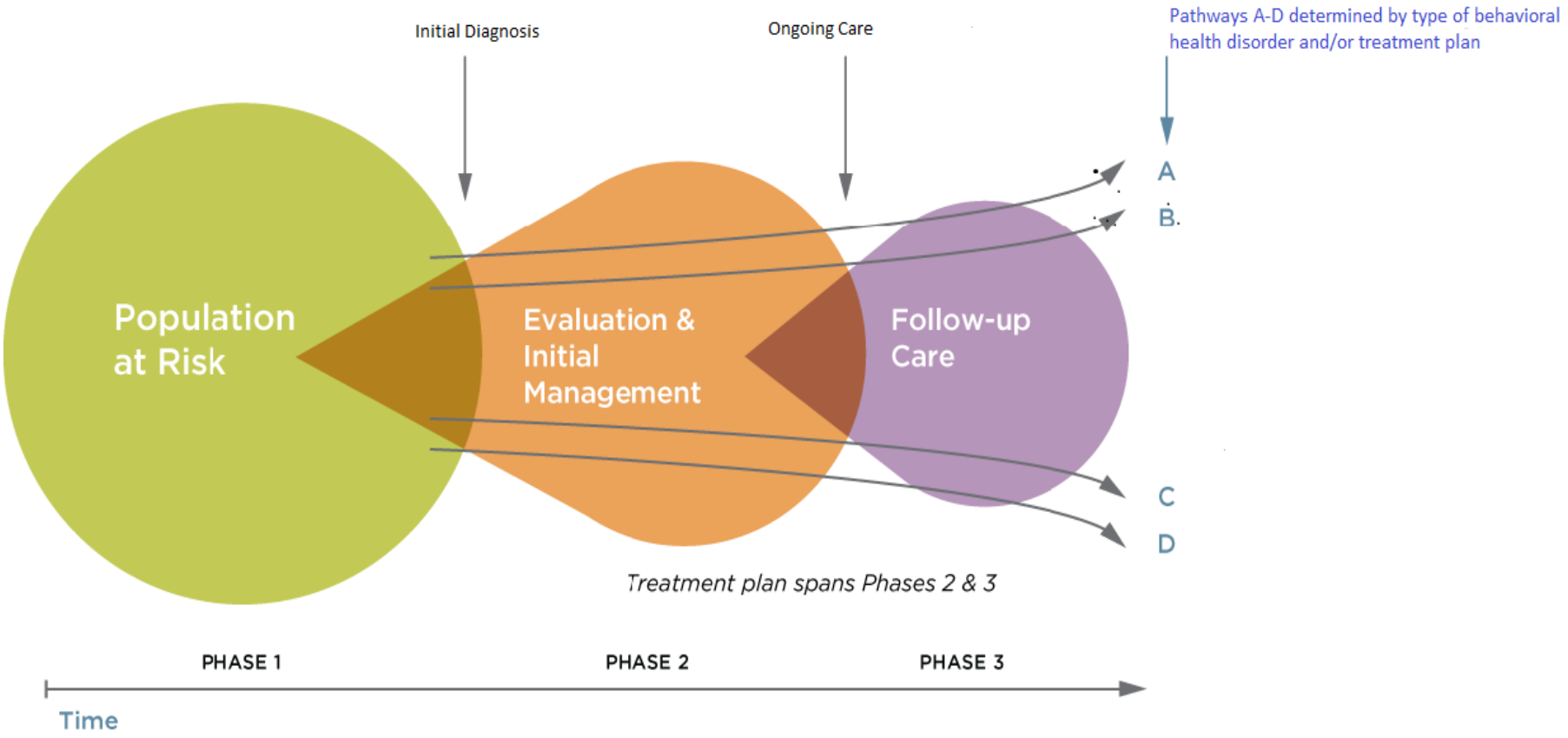
- *Respond to NQF member or public queries about the project*
- *Maintain documentation of project activities*
- *Post project information to NQF website*
- *Work with measure developers to provide necessary information and communication for the SC to fairly and adequately evaluate measures for endorsement*
- *Publish final project report*

Questions?

Overview of NQF's Behavioral Health Portfolio

Behavioral Portfolio of Measures

- This project will evaluate measures related to behavioral health conditions that can be used for accountability and public reporting for all populations and in all settings of care. The fourth phase of this project will address topic areas including:
  - *Alcohol and substance use*
  - *Tobacco use*
  - *Attention deficit/hyperactivity disorder*
  - *Depression*
- NQF solicits new measures for possible endorsement
- NQF currently has more than 50 endorsed measures within the area of behavioral health. Endorsed measures undergo periodic evaluation to maintain endorsement – "maintenance".


Initial Diagnosis
Ongoing Care
Pathways A-D determined by type of behavioral health disorder and/or treatment plan
Population at Risk
Evaluation & Initial Management
Follow-up Care
A
B
C
D
Treatment plan spans Phases 2 & 3
PHASE 1
PHASE 2
PHASE 3
Time

Behavioral Health Portfolio of NQF-endorsed measures

*Measures for maintenance evaluation

- 0008: Experience of Care and Health Outcomes (ECHO) Survey (behavioral health, managed care versions) (CMS)*
- 0027: Medical Assistance With Smoking and Tobacco Use Cessation (NCQA)*
- 0028: Preventive Care & Screening: Tobacco Use: Screening & Cessation Intervention (PCPI Foundation)*
- 3185: Preventive Care and Screening: Tobacco Use: Screening and Cessation Intervention (eMeasure) (PCPI Foundation)
- 0108: Follow-Up Care for Children Prescribed ADHD Medication (NCQA)*
- 0576: Follow-Up After Hospitalization for Mental Illness (NCQA)*
- 3132: Preventive Care and Screening: Screening for Clinical Depression and Follow-Up Plan (eMeasure) (CMS)
- 3148: Preventive Care and Screening: Screening for Clinical Depression and Follow-Up Plan (CMS)*
- 3172: Continuity of Pharmacotherapy for Alcohol Use Disorder (RAND Corporation)
- 3175: Continuity of Pharmacotherapy for Opioid Use Disorder (RAND Corporation)

Activities and Timeline

*All times ET

Meeting	Date/Time
Standing Committee Orientation Call	January 19, 2017 1:00 PM – 3:00 PM EST
Measure Evaluation Q&A Webinar	February 16, 2017 1:00 PM – 3:00 PM EST
Committee Measure Evaluation Surveys Due	February 20, 2017
In-Person Meeting (2 days in Washington, DC)	February 28 & March 1, 2017
Post Meeting Webinar	March 9, 2017 1:00 PM- 3:00 PM EST
Draft Reporting Commenting Period	April 5-May 4, 2017
Post Draft Report Comment Webinar	TBD
Member Vote	June 5-19, 2017
Appeals	July 14-August 14, 2017

Questions?

Measure Evaluation Criteria Overview

NQF Measure Evaluation Criteria for Endorsement

NQF endorses measures for accountability applications (public reporting, payment programs, accreditation, etc.) as well as quality improvement.

- Standardized evaluation criteria
- Criteria have evolved over time in response to stakeholder feedback
- The quality measurement enterprise is constantly growing and evolving – greater experience, lessons learned, expanding demands for measures – the criteria evolve to reflect the ongoing needs of stakeholders

Major Endorsement Criteria
Hierarchy and Rationale (page 31)

- ***Importance to measure and report**: Goal is to measure those aspects with greatest potential of driving improvements; if not important, the other criteria are less meaningful (**must-pass**)*
- ***Reliability and Validity-scientific acceptability of measure properties** : Goal is to make valid conclusions about quality; if not reliable and valid, there is risk of improper interpretation (**must-pass**)*
- ***Feasibility**: Goal is to, ideally, cause as little burden as possible; if not feasible, consider alternative approaches*
- ***Usability and Use**: Goal is to use for decisions related to accountability and improvement; if not useful, probably do not care if feasible*
- ***Comparison to related or competing measures***

Criterion #1: Importance to Measure and Report (page 33-41)

1. **Importance to measure and report -** Extent to which the specific measure focus is evidence-based and important to making significant gains in healthcare quality where there is variation in or overall less-than-optimal performance.

   1a. Evidence: *the measure focus is evidence-based (page 34-39)*

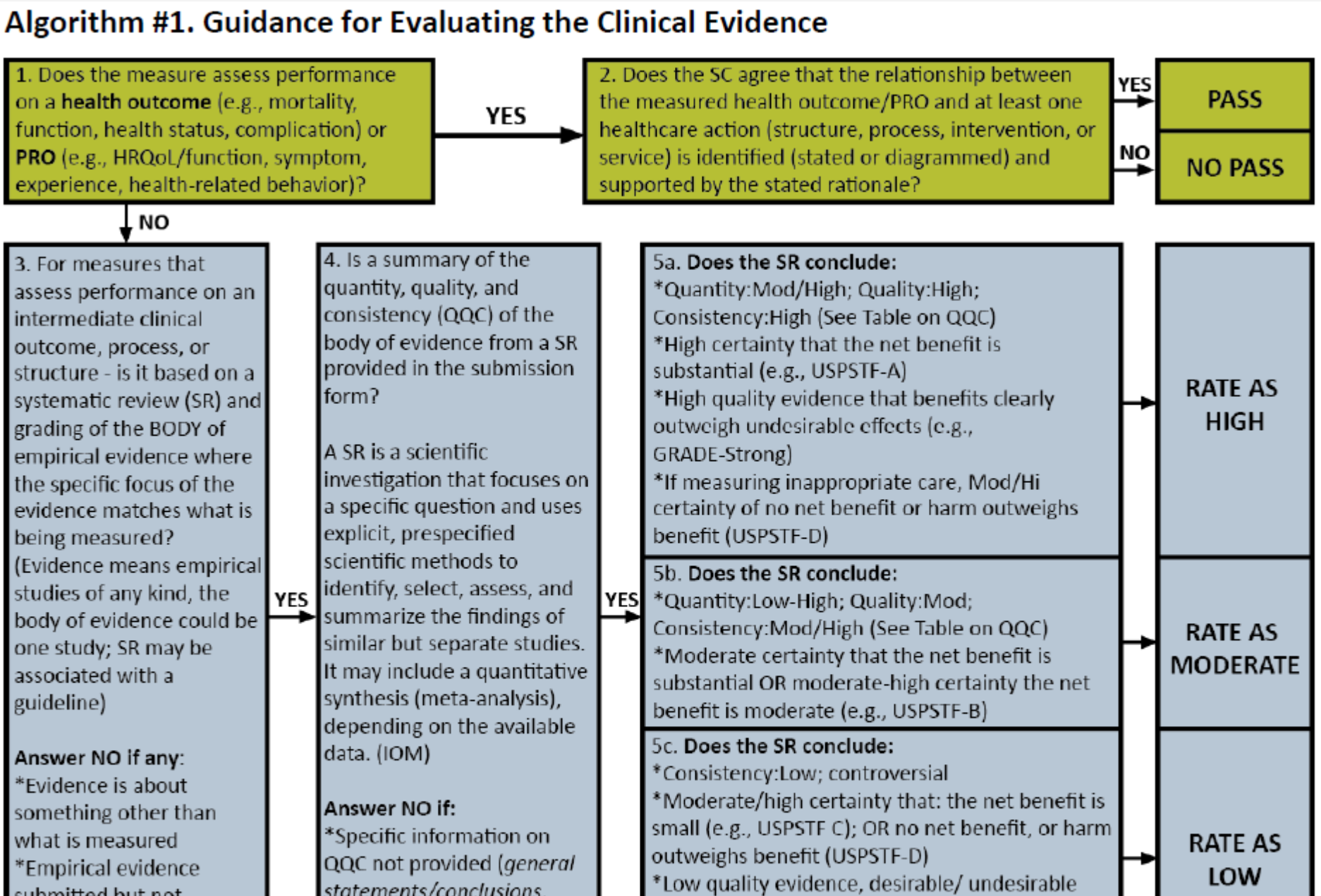
   1b. Opportunity for Improvement: *demonstration of quality problems and opportunity for improvement, i.e., data demonstrating considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or disparities in care across population groups (pages 39-40)*

   1c. Quality construct and rationale *(composite measures only)*

Subcriteron 1a: Evidence (page 34-39)

- Outcome measures
  - *A rationale (which often includes evidence) for how the outcome is influenced by healthcare processes or structures.*
- Structure, process, intermediate outcome measures
  - *The quantity, quality, and consistency of the body of evidence underlying the measure should demonstrate that the measure focuses on those aspects of care known to influence desired patient outcomes*
    - » Empirical studies (expert opinion is not evidence)
    - » Systematic review and grading of evidence
      - *Clinical Practice Guidelines – variable in approach to evidence review*

Rating Evidence: Algorithm #1 – page 36



Criterion #1: Importance to measure and report Criteria emphasis is different for new vs. maintenance measures

New measures	Maintenance measures
• Evidence – Quantity, quality, consistency (QQC) • Established link for process measures with outcomes	**DECREASED EMPHASIS**: Require measure developer to attest evidence is unchanged evidence from last evaluation; Standing Committee to affirm no change in evidence IF changes in evidence, the Committee will evaluate as for new measures
• Gap – opportunity for improvement, variation, quality of care across providers	**INCREASED EMPHASIS**: data on current performance, gap in care and variation

Criterion #2: Reliability and Validity– Scientific Acceptability of Measure Properties (page 41 - 51)

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of health care delivery

2a. **Reliability (must-pass)**

2a1. Precise specifications including exclusions

2a2. Reliability testing—data elements or measure score

2b. **Validity (must-pass)**

2b1. Specifications consistent with evidence

2b2. Validity testing—data elements or measure score

2b3. Justification of exclusions—relates to evidence

2b4. Risk adjustment—typically for outcome/cost/resource use

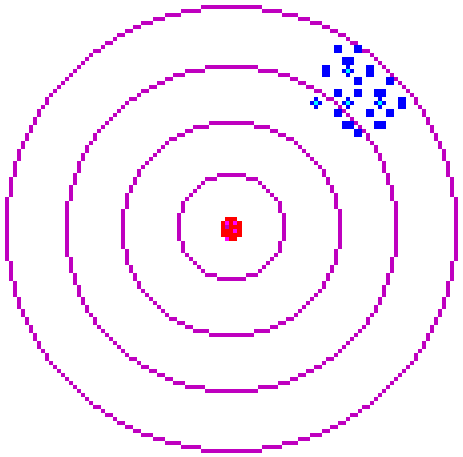
2b5. Identification of differences in performance

2b6. Comparability of data sources/methods

2b7. Missing data

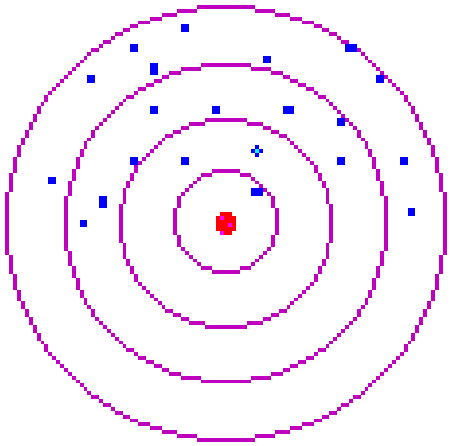
Reliability and Validity (page 42)

Assume the center of the target is the true score…



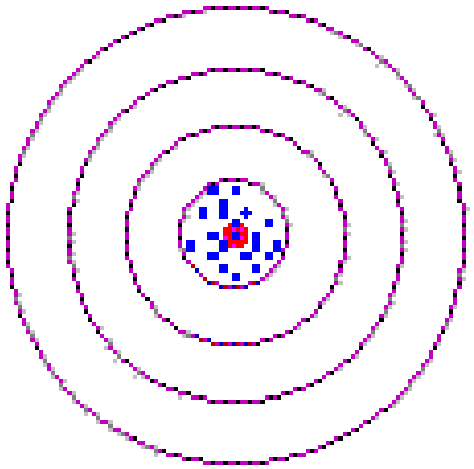
Reliable Not Valid

Consistent, but wrong



Neither Reliable Nor Valid

Inconsistent & wrong



Both Reliable And Valid

Consistent & correct

Measure Testing – Key Points (page 43)

Empirical analysis to demonstrate the reliability and validity of the *measure as specified,* including analysis of issues that pose threats to the validity of conclusions about quality of care such as exclusions, risk adjustment/stratification for outcome and resource use measures, methods to identify differences in performance, and comparability of data sources/methods.

Reliability Testing (page 43)
Key points - page 44

- Reliability of the ***measure score*** refers to the proportion of variation in the performance scores due to systematic differences across the measured entities in relation to random variation or noise (i.e., the precision of the measure).
  - *Example - Statistical analysis of sources of variation in performance measure scores (signal-to-noise analysis)*

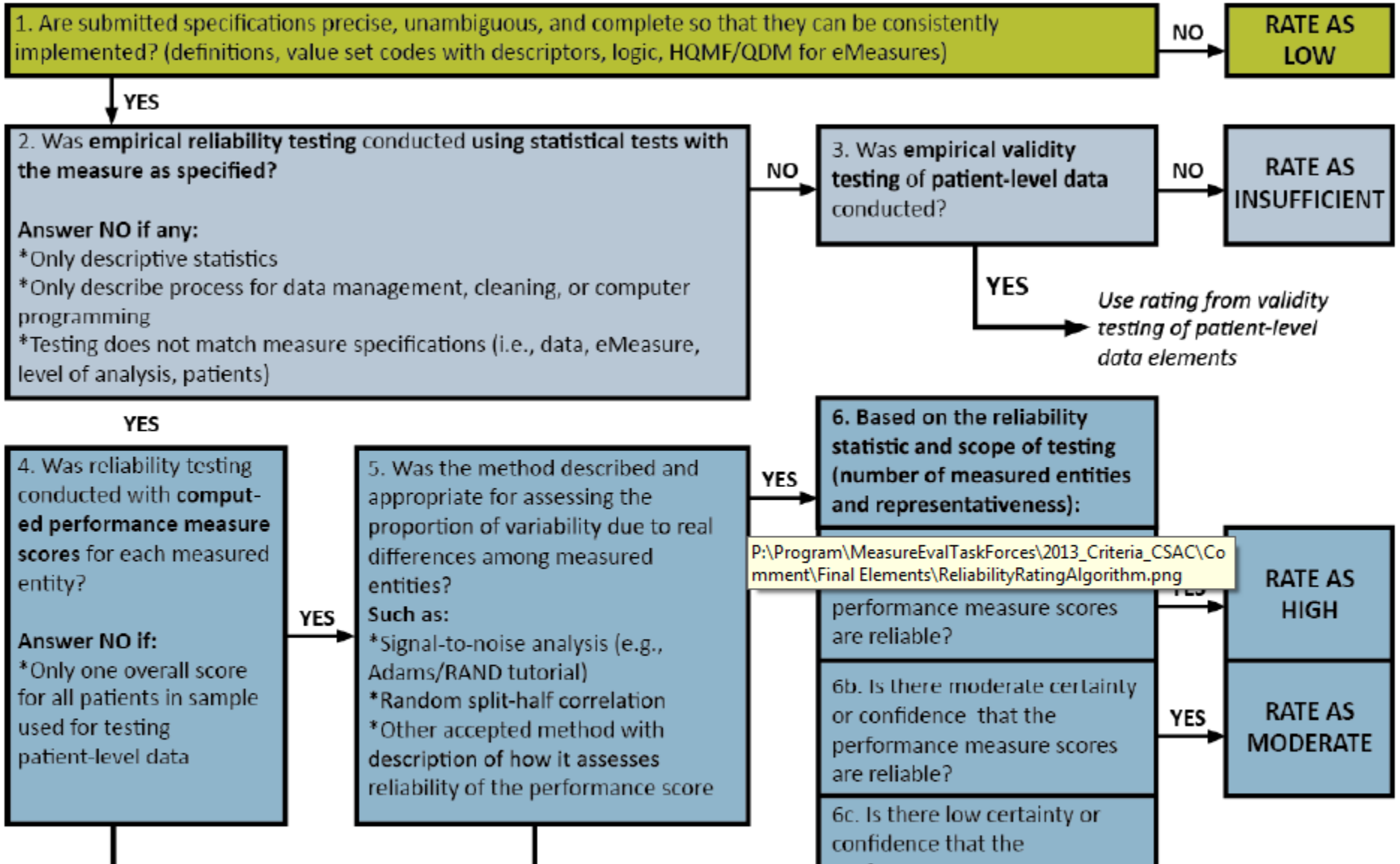
- Reliability of the ***data elements*** refers to the repeatability/reproducibility of the data and uses patient-level data
  - *Example –inter-rater reliability*

- Consider whether testing used an appropriate method and included adequate representation of providers and patients and whether results are within acceptable norms

- Algorithm #2 – page 45

Rating Reliability: Algorithm #2 – page 45

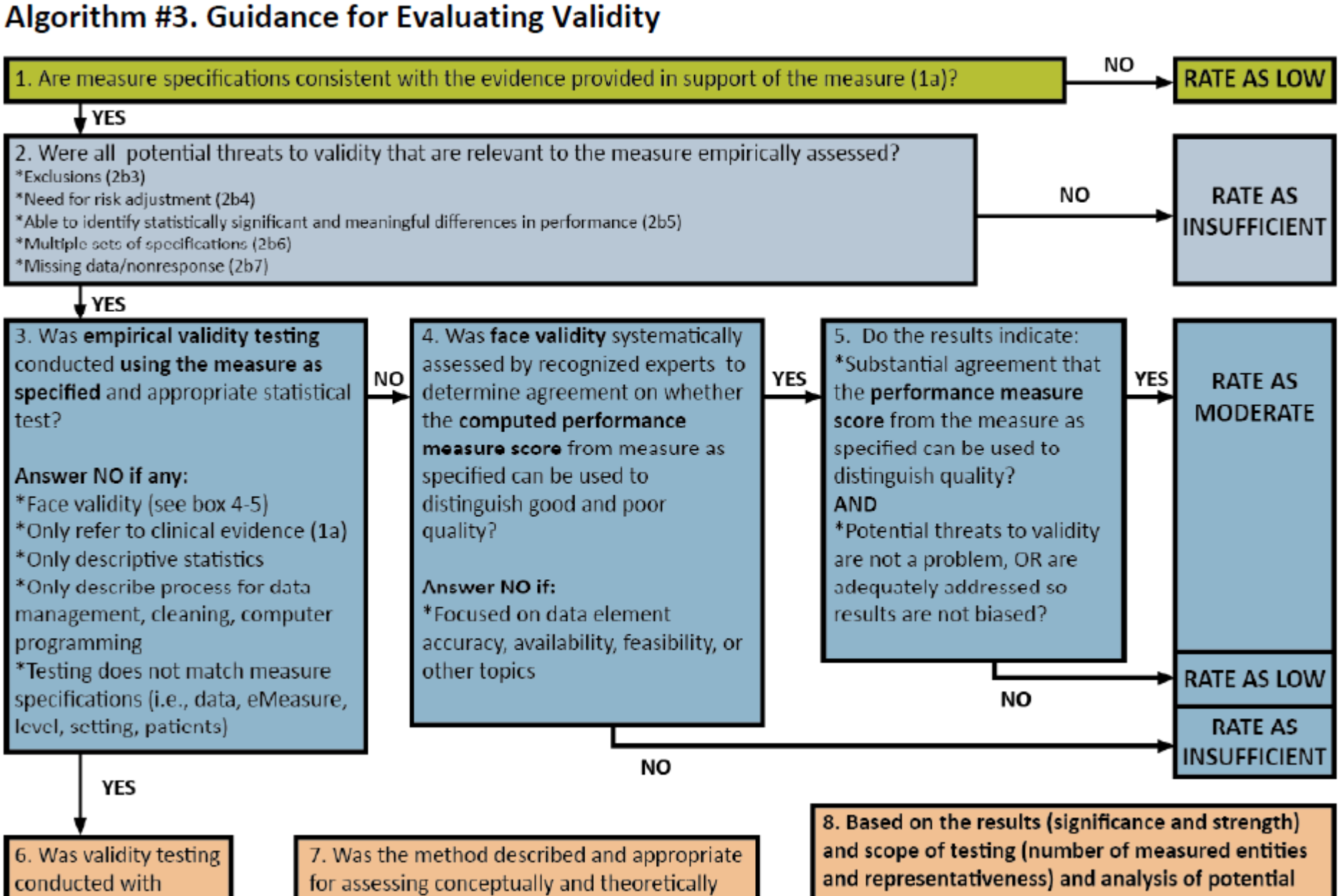
Algorithm #2. Guidance for Evaluating Reliability



Validity testing (pages 46 - 51)
Key points – page 49

- **Empirical testing**
  - *Measure score* – assesses a hypothesized relationship of the measure results to some other concept; assesses the correctness of conclusions about quality
  - *Data element* – assesses the correctness of the data elements compared to a "gold standard"
- **Face validity**
  - *Subjective determination by experts that the measure appears to reflect quality of care*

Rating Validity: Algorithm #3 – page 50



Threats to Validity

- Conceptual
  - *Measure focus is not a relevant outcome of healthcare or not strongly linked to a relevant outcome*
- Unreliability
  - *Generally, an unreliable measure cannot be valid*
- Patients inappropriately excluded from measurement
- Differences in patient mix for outcome and resource use measures
- Measure scores that are generated with multiple data sources/methods
- Systematic missing or “incorrect” data (unintentional or intentional)

Criterion #2: Scientific Acceptability

New measures	Maintenance measures
• Measure specifications are precise with all information needed to implement the measure	NO DIFFERENCE: Require updated specifications
• Reliability • Validity (including risk-adjustment)	**DECREASED EMPHASIS**: If prior testing adequate, no need for additional testing at maintenance with certain exceptions (e.g., change in data source, level of analysis, or setting) Must address the questions for SDS Trial Period

Criterion #3: Feasibility (page 51)
Key Points – page 52

Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement.

3a: Clinical data generated during care process

3b: Electronic sources

3c: Data collection strategy can be implemented

Criterion #4: Usability and Use (page 52)
Key Points – page 53

Extent to which potential audiences (e.g., consumers, purchasers, providers, policymakers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a: Accountability and Transparency: Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement

4b: Improvement: Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated

4c: Benefits outweigh the harms: The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4d: Vetting by those being measured and others: Those being measured have been given results and assistance in interpreting results; those being measured and others have been given opportunity for feedback; the feedback has been considered by developers.

Criteria #3-4: Feasibility and Usability and Use

New measures	Maintenance measures
Feasibility	
• Measure feasible, including eMeasure feasibility assessment	NO DIFFERENCE: Implementation issues may be more prominent
Usability and Use	
• Use: used in accountability applications and public reporting	**INCREASED EMPHASIS**: Much greater focus on measure use and usefulness, including both impact and unintended consequences
• Usability: impact and unintended consequences	

Criterion #5: Related or Competing Measures (page 53-54)

If a measure meets the four criteria and there are endorsed/new related measures (same measure focus or same target population) or competing measures (both the same measure focus and same target population), the measures are compared to address harmonization and/or selection of the best measure.

- 5a. The measure specifications are harmonized with related measures **OR** the differences in specifications are justified.
- 5b. The measure is superior to competing measures (e.g., is a more valid or efficient way to measure) **OR** multiple measures are justified.

Evaluation process

- **Preliminary analysis**: To assist the Committee evaluation of each measure against the criteria, NQF staff will prepare a preliminary analysis of the measure submission and offer preliminary ratings for each of the criteria.
  - *These will be used as a starting point for the Committee discussion and evaluation*
- **Individual evaluation assignments:** Each Committee member will be assigned a subset of measures for in-depth evaluation.
  - *Those who are assigned measures will lead the discussion of their measures with the entire Committee*

Evaluation process (continued)

- **Measure evaluation and recommendations at the in-person meeting:** The entire Committee will discuss and rate each measure against the evaluation criteria and make recommendations for endorsement.

Recommendation for Endorsement and Endorsement +

- The Committee votes on whether to recommend a measure for NQF endorsement.
- Staff will inform the Committee when a measure has met the criteria for possible "Endorsement +" designation:
  - *Meets evidence criteria without exception*
  - *Good results on reliability testing of the measure score*
  - *Good results on empirical validity testing of the measure score (not just face validity)*
  - *Well-vetted in real world settings by those being measured and others*
- Committee votes on recommending the "Endorsement +" designation, indicating that the measure exceeds NQF criteria in key areas.

Questions?

SDS Trial Period Overview

Background

- During a **two-year trial period,** adjustment of measures for socio-demographic (SDS) factors **will no longer be prohibited**
- Each measure must be assessed individually to determine if SDS adjustment is appropriate (included as part of validity subcriterion)
- The Standing Committee will continue to evaluate the measure as a whole, including the appropriateness of the risk adjustment approach used by the measure developer
- Efforts to implement SDS adjustment may be constrained by data limitations and data collection burden

Standing Committee Evaluation

- The Standing Committee will be asked to consider the following questions:
  - *Is there a conceptual relationship between the SDS factor and the measure focus?*
  - *What are the patient-level sociodemographic variables that were available and analyzed during measure development?*
  - *Does empirical analysis (as provided by the measure developer) show that the SDS factor has a significant and unique effect on the outcome in question?*
  - *Does the reliability and validity testing match the final measure specifications?*

Questions?

SharePoint Overview

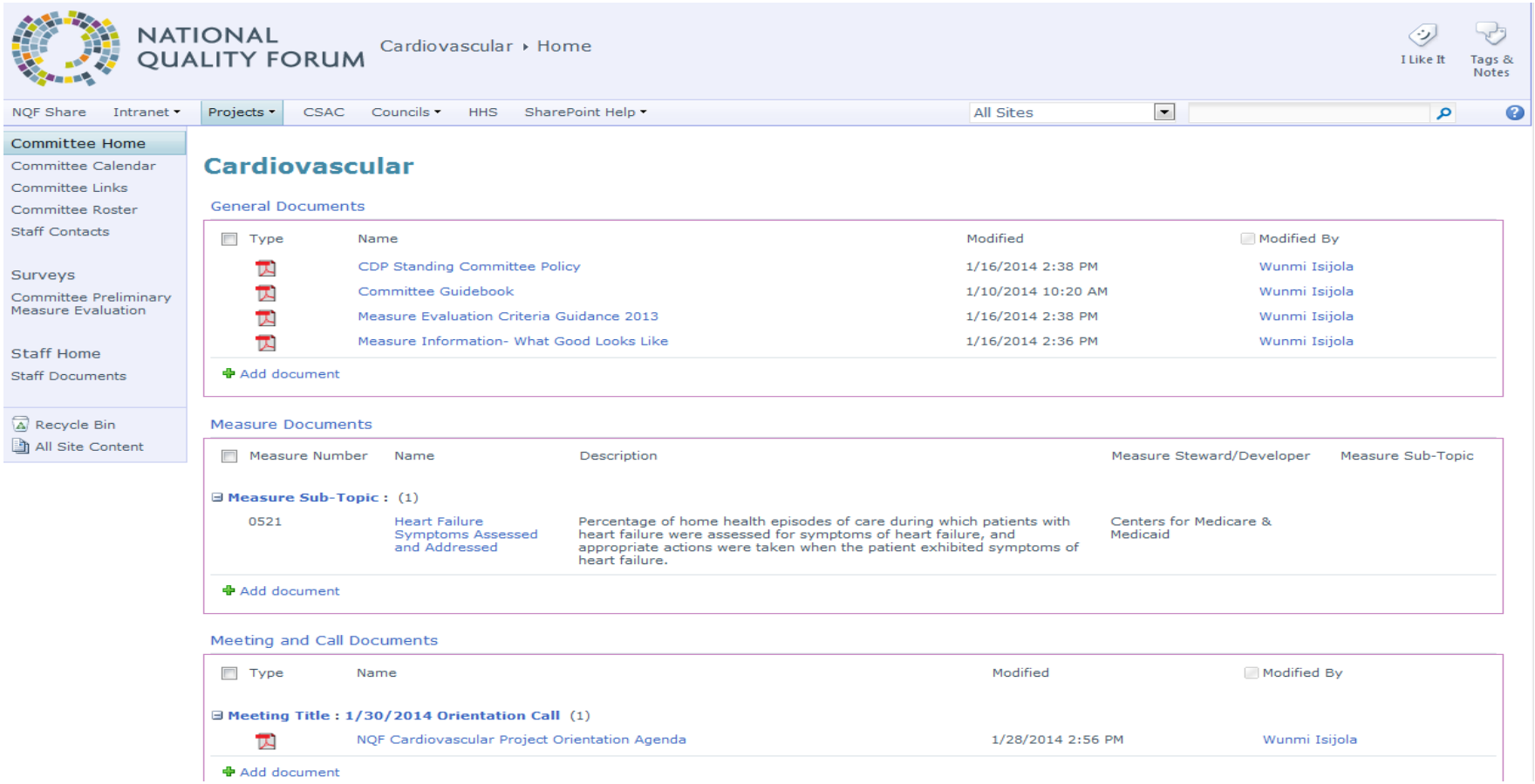
SharePoint Overview

http://share.qualityforum.org/Projects/behavioralhealth/SitePages/Home.aspx

- Accessing SharePoint
- Standing Committee Policy
- Standing Committee Guidebook
- Measure Document Sets
- Meeting and Call Documents
- Committee Roster and Biographies
- Calendar of Meetings

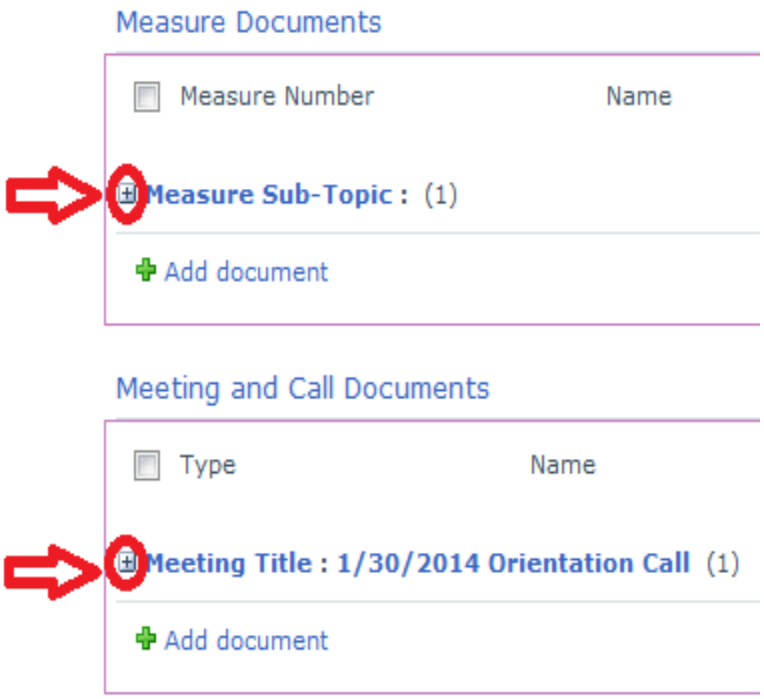
SharePoint Overview

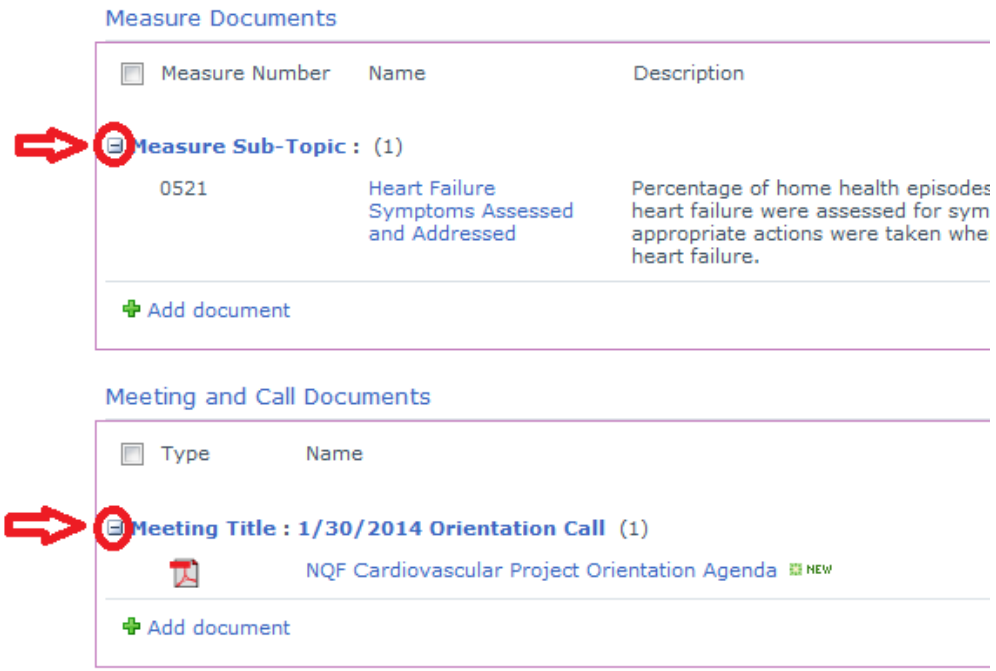
- **Screen shot of homepage:**



SharePoint Overview

- Please keep in mind:





Measure Worksheet and Measure Information

- Measure Worksheet
  - *Preliminary analysis, including eMeasure Technical Review if needed, and preliminary ratings*
  - *Pre-evaluation comments*
  - *Public comments*
  - *Information submitted by the developer*
    - » Evidence and testing attachments
    - » Spreadsheets
    - » Additional documents

Next Steps

- Measure Evaluation Q&A Call
  - *February 16, 2017, 1:00-3:00 PM ET*
- In-Person Meeting
  - *February 28-March 1, 2017*
- Post Meeting Webinar
  - *March 9, 2017, 1:00-3:00 PM ET*
- Post Comment Webinar
  - *TBD*

Project Contact Info

- Email: behavioralhealth@qualityforum.org
- NQF Phone: 202-783-1300
- Project page:
http://www.qualityforum.org/Project_Pages/behavioralhealth.aspx
- SharePoint site:
http://share.qualityforum.org/Projects/behavioralhealth/SitePages/Home.aspx

Questions?